

On Optimal Rules of Persuasion

Jacob Glazer

Faculty of Management, Tel Aviv University

and

Department of Economics, Boston University

glazer@post.tau.ac.il

and

Ariel Rubinstein

School of Economics, Tel Aviv University

rariel@post.tau.ac.il

Before reading the paper we advise you to play on-line our “Persuasion game”:

<http://gametheory.tau.ac.il/exp5/>

first version: 1 May 2003

revised version: 3 June 2003

Abstract

A speaker attempts to persuade a listener to accept a certain request. The listener is not sure that the request is justified. The conditions under which the listener should accept the request depend on the values of two aspects known only to the speaker. In order to persuade the listener, the speaker is able to send him a message. The listener can then check the true value of at most one of the two aspects (which is probably to be chosen randomly) after which he either accepts or denies the request. The paper models such a situation and studies the persuasion rules that minimize the probability of the listener making a mistake.

Key words: persuasion, mechanism design, hard evidence, debates.

Classifications: C610, C72, D78, B201, R316.

1. Introduction

In daily life we commonly encounter situations in which one agent (*the speaker*) tries to persuade another agent (*the listener*) to take a certain action by arguing that according to the information he possesses this action is mutually beneficial. In this paper we study a simple model that captures this situation.

In our model the listener has to choose between two actions a and r . The listener's preferred action critically depends on the speaker's type - a realization of two independent random variables (aspects) - initially known only to the speaker, whereas the speaker would like the listener to choose the action a regardless of his type.

We study a family of mechanisms in which the speaker sends a message to the listener about his type and the listener can then choose to ascertain the true realization of at most one of the two aspects that characterize the speaker's type. On the basis of the speaker's claims and the acquired "hard" evidence, the listener is either persuaded to take the speaker's proposed action a or not.

Following are some real life examples in the spirit of this model:

A worker wishes to be hired by an employer for a certain position. The worker tells the employer about his previous experience in two similar jobs. The employer will hire the worker if his average performance in his two previous jobs was above a certain minimal level. However, before making the final decision the employer has sufficient time to thoroughly interview at most one of the candidate's previous employers.

A suspect is arrested on the basis of testimonies provided by two witnesses. The suspect's lawyer claims that their testimonies to the police have serious inconsistencies and therefore his client should be released. The judge's preferred decision rule is to release the suspect only if the two testimonies substantially contradict one another; however, he is able to investigate at most one of the two witnesses.

A doctor claims that he has correctly used two procedures to treat a patient who suffers from

two chronic illnesses. An investigator in the case is asked to determine whether the combination of the two procedures was harmful. The investigator has access to the doctor's full report but verifying the details of more than one procedure is too costly.

A decision maker asks a consultant for advice. The decision maker knows that the consultant is better informed about the state of the world than he is, but he also knows that the consultant sometimes has different interests in recommending which action should be taken. The decision maker can listen to the consultant's advice but can only verify some of the facts that the consultant claims to be true.

We are interested in the properties of the mechanisms that are optimal from the point of view of the listener, namely, those in which it is least likely that the listener will choose the wrong action. A mechanism is composed of several elements: a set of messages the speaker can choose from; a function that specifies which aspect is to be checked depending on the speaker's message; and the action the listener finally takes as a function of the message sent by the speaker and the acquired information. In other words, a mechanism specifies the conditions under which the listener is persuaded by the speaker. In our scenario, the listener does not have tools to deter the speaker from cheating and thus we can expect that the speaker will always argue that his information indicates that the action a should be taken. The problem therefore is to decide which rules the listener should follow in order to minimize the probability of making a mistake.

Two types of mechanisms will serve a special role in our analysis:

A) Deterministic mechanisms – for each of the two aspects certain criteria are determined and the speaker's preferred action is chosen if he can show that his type meets these prespecified criteria in at least one of the two aspects. In the first example above, a deterministic mechanism would be equivalent to asking the worker to provide a reference from one of his two previous employers which meets certain criteria.

B) Random mechanisms – the speaker is asked to report his type; one aspect is then chosen randomly to be checked; and the action a is taken if and only if the speaker's report is not

refuted. Returning to the first example above, a random mechanism would involve first asking the worker to justify his application by reporting his performance in each of his previous two jobs. Based on his report, the employer then randomly selects one of the two previous employers to interview and accepts the applicant if his report is not refuted.

The main results of the paper are as follows:

1) Finding the optimal mechanism is equivalent to solving an auxiliary linear programming problem in which the objective function minimizes the number of mistakes. The inequality constraints are derived from a condition that we call the *L*-principle which can be demonstrated using the first example above: Assume that the employer wishes to hire the worker only if his performance in the two previous jobs was above a certain level. Further assume that the worker's performances in each job can be evaluated and assigned to one of two categories: good or bad. Since the worker who has done well in only one of the jobs can claim that he performed well in both and since the employer can verify only one reference, the sum of probabilities that a mistaken action will be taken regarding the three types is at least one.

2) An optimal mechanism with a very simple structure always exists: the speaker is asked to report his type; if according to this report the action r should be taken, then the listener takes it; otherwise the listener tosses a fair coin and depending on its outcome he either takes the action a or r , or he checks one of the aspects and takes the action a if the speaker's claim regarding this aspect is confirmed.

3) Under certain "convexity" and "monotonicity" conditions on the set of the speaker's types under which the action r should be taken, there exists an optimal mechanism that is deterministic.

2. The Model

Let $\{1, \dots, n\}$ be a set of random variables which we call *aspects*. Most of the analysis will be conducted for $n = 2$. The realization of aspect k is a member of a set X_k . There are two agents: the *speaker* and the *listener*. A *problem* is a pair (X, A) , where $A \subseteq X = \times_{k=1, \dots, n} X_k$. A problem is *finite* if the set X is finite. We attach to X a uniform probability measure. A member of X is called a (speaker's) *type* and is interpreted as a possible characterization of the speaker.

The listener has to take one of two actions: a (accept) or r (reject). The listener is interested in taking the action a if the speaker's type is in A and the action r if the type is in $R = X - A$. The speaker, regardless of his type, prefers the listener to take the action a .

The speaker knows his type while the listener only knows its distribution. The listener can *check*, that is, ascertain the realization of, at most one of the n aspects.

Let Q be the set of all lotteries $\langle \pi_0, d_0; \pi_1, d_1; \dots; \pi_n, d_n \rangle$ where $(\pi_i)_{i=0,1,\dots,n}$ is a probability vector and $d_k : X_k \rightarrow \{a, r\}$ where $X_0 = \{e\}$ is an arbitrary singleton set (that is d_0 is a constant). An element in Q is interpreted as a possible response of the listener to a message. With probability π_0 no aspect is checked and the action $d_0 \in \{a, r\}$ is taken and with probability π_k ($k = 1, \dots, n$) aspect k is checked and if its realization is x_k the action $d_k(x_k)$ is taken.

A *mechanism* is (M, f) , where M is a set (of *messages*) and $f : M \rightarrow Q$. A *direct mechanism* is one where $M = X$. For a direct mechanism (X, f) we say that following a message m the *mechanism verifies aspect k with probability π_k* if $f(m) = \langle \pi_0, d_0; \pi_1, d_1; \dots; \pi_n, d_n \rangle$ is such that $d_k(x_k) = a$ iff $x_k = m_k$. The *fair random mechanism* is the direct mechanism by which, for every $m \in A$, the speaker verifies each aspect with probability $1/n$ and, for every $m \in R$, he chooses the action r . A mechanism is *deterministic* if for every $m \in M$ the lottery $f(m)$ is degenerate (for some k , $\pi_k = 1$).

For every lottery $q = \langle \pi_0, d_0; \pi_1, d_1; \dots; \pi_n, d_n \rangle$ and every type x define $q(x)$ to be the probability that the action a is taken when the lottery q is applied to type x , that is,

$$q(x) = \sum_{d_k(x_k)=a} \pi_k.$$

Given a mechanism (M, f) let μ_x be the probability that the listener takes the

wrong action with respect to type x , assuming that the speaker sends a message that maximizes the probability that the listener takes the action a . Formally, for $x \in R$ let $\mu_x = \max_{m \in M} f(m)(x)$ and for $x \in A$ let $\mu_x = 1 - \max_{m \in M} f(m)(x)$. The *mistake probability* induced by the mechanism is $\int_{x \in X} \mu_x$. For a finite problem the mistake probability is $\sum_{x \in X} \mu_x / |X|$ and we refer to $\sum_{x \in X} \mu_x$ as the *number of mistakes*.

Given a problem (X, A) , an *optimal mechanism* is one that minimizes the mistake probability. Applying the revelation principle we focus on direct mechanisms only.

Following is a concrete example.

Example 1: Let $X_1 = X_2 = [0, 1]$ and let $A = \{(x_1, x_2) | x_1 + x_2 \geq 1\}$.

If the speaker is not asked to talk, the lowest probability of mistake is $1/4$. This mistake probability can be obtained by a mechanism in which aspect 1 is checked with probability 1 and the action a is taken iff aspect 1's value is at least $1/2$ (formally, $M = \{e\}$ and $f(e)$ is the degenerate lottery where $\pi_1 = 1$ and $d_1(x_1) = a$ iff $x_1 \geq 1/2$).

In this example, letting the speaker talk can improve matters. Consider the following deterministic direct mechanism ($M = X$) characterized by two numbers z_1 and z_2 . Following the receipt of a message (m_1, m_2) , the speaker verifies aspect 1 if $m_1 \geq z_1$ and verifies aspect 2 if $m_1 < z_1$ but $m_2 \geq z_2$. If $m_k < z_k$ for both k the action r is taken. One interpretation of this mechanism is that in order to persuade the listener, the speaker has to show that the realization of at least one of the aspects is above some threshold (which may be different for each aspect). The set of types for which the outcome will be wrong consists of the three shaded triangles shown in figure 1a:

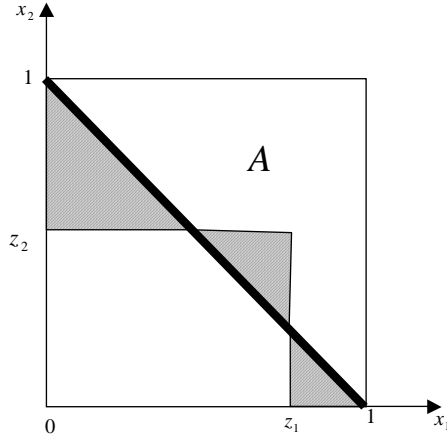


Figure 1a

One can see that the optimal thresholds are $z_1 = z_2 = 2/3$ yielding a mistake probability of $1/6$. Is it possible to obtain a lower probability of mistake by applying a randomized mechanism by which each of the two aspects is checked with some positive probability? We will return to this question later.

3. A Basic Proposition

From now on, assume $n = 2$. For simplicity of notation we write μ_{ij} for $\mu_{(i,j)}$. The following proposition is key to our analysis (note that it is valid for both finite and infinite X):

Proposition 0: (The L principle) Let (X,A) be a problem. For any mechanism and for any three elements $(i,j) \in A$, $(i,s) \in R$ and $(t,j) \in R$ it must be that $\mu_{ij} + \mu_{is} + \mu_{tj} \geq 1$.

Proof: Let $f(i,j) = \langle \pi_0, d_0; \pi_1, d_1; \pi_2, d_2 \rangle$. Let δ_p be 1 if p is true and 0 otherwise. Then

$$\mu_{ij} = \pi_0 \delta_{d_0=r} + \pi_1 \delta_{d_1(i)=r} + \pi_2 \delta_{d_2(j)=r}$$

$$\mu_{is} \geq \pi_0 \delta_{d_0=a} + \pi_1 \delta_{d_1(i)=a}$$

$$\mu_{tj} \geq \pi_0 \delta_{d_0=a} + \pi_2 \delta_{d_2(j)=a}.$$

Therefore

$$\mu_{ij} + \mu_{is} + \mu_{tj} \geq \pi_0 \delta_{d_0=r} + \pi_1 \delta_{d_1(i)=r} + \pi_2 \delta_{d_2(j)=r} + \pi_0 \delta_{d_0=a} + \pi_1 \delta_{d_1(i)=a} + \pi_0 \delta_{d_0=a} + \pi_1 \delta_{d_2(j)=a} = 1 + \pi_0 \delta_{d_0=a} \geq 1.$$

■

The idea of the proof is as follows: whatever the outcome of the randomization following a message m sent by type $(i, j) \in A$, either the mistaken action r is taken, or at least one of the two types (i, s) and (t, j) in R can induce the wrong action a by sending m .

We define an L to be any set of three elements $(i, j) \in A$, $(i, s) \in R$ and $(t, j) \in R$. We refer to the result of Proposition 0 (the sum of mistakes in every L is at least 1) as *the L-principle*. Obviously, the L -principle extends to the case of $n > 2$.

Example 1 (again): To demonstrate the usefulness of Proposition 0, let us go back to Example 1 and show that the optimal mechanism is deterministic. We have already shown that there is a deterministic mechanism with mistake probability of $1/6$. To see that the mistake probability of any mechanism is at least $1/6$, divide the unit square into 9 equal squares and divide each square into two triangles as shown in Figure 1b. The set $M_1 = \{(x_1, x_2) \in A \mid x_1 \leq 2/3 \text{ and } x_2 \leq 2/3\}$ is one of the three triangles denoted in the figure by the number 1.

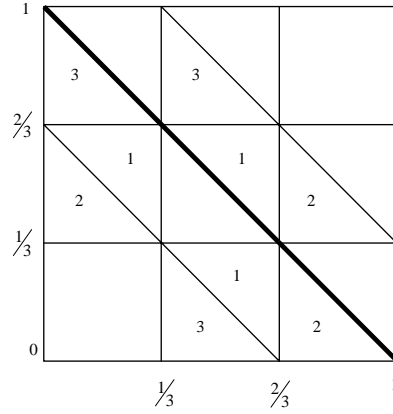


Figure 1b

Any three points $x = (x_1, x_2) \in M_1$, $y = (x_1 - 1/3, x_2) \in R$ and $z = (x_1, x_2 - 1/3) \in R$ establish an L . By Proposition 0, $\mu_x + \mu_y + \mu_z \geq 1$. The collection of all these L 's is a set of disjoint sets whose union is the three triangles denoted in the figure by the number 1. Therefore the integral of μ over these three triangles must be at least the size of M_1 , namely $1/18$. Similar considerations regarding the three triangles denoted by the number 2 and the three triangles denoted by the number 3 imply that $\int_{x \in X} \mu_x \geq 1/6$.

4. An Equivalent Linear Programming Problem

We will now show that finding the optimal mechanism is equivalent to solving an auxiliary linear programming problem.

Let (X, A) be a finite problem. Define $P(X, A)$ to be the linear programming problem:

$$\min \sum_{x \in X} \mu_x$$

$$\text{s.t. } \mu_{ij} + \mu_{is} + \mu_{tj} \geq 1 \text{ for all } (i, j) \in A, (i, s) \in R \text{ and } (t, j) \in R$$

and $0 \leq \mu_x$ for all $x \in X$

We will show that the solution to $P(X,A)$ coincides with the vector of mistake probabilities induced by an optimal mechanism for the problem (X,A) .

Note that not every vector which satisfies the constraints of $P(X,A)$ can be induced by a mechanism. Even if we add the constraints $\mu_x \leq 1$, “incentive compatibility” implies additional constraints on the vectors of mistake probabilities induced by a mechanism. For example, if $(i,j) \in A$, $(i,s) \in A$ and $(t,j) \in A$ then it is impossible that $\mu_{ij} = 0$ while both $\mu_{is} = 1$ and $\mu_{tj} = 1$, since at least one of the types, (i,s) or (t,j) , will pretend to be (i,j) , increasing the probability that the action taken is a .

Proposition 1: Let $(\mu_x)_{x \in X}$ be a solution to $P(X,A)$. Then, there is an optimal mechanism for (X,A) such that for all $x \in X$ the probability of mistake with respect to type x is μ_x .

Proof:

Step 1: Note that for every $x \in X$ it must be that $\mu_x \leq 1$ and either $\mu_x = 0$ or there are two other elements y and z such that $\{x,y,z\}$ establish an L and $\mu_x + \mu_y + \mu_z = 1$. Otherwise we could reduce μ_x and stay within the constraints.

Step 2: By Proposition 0 any vector of mistake probabilities induced by a mechanism satisfies the constraints. Thus, it is sufficient to construct a mechanism such that the mistake probability with respect to any $x \in X$ is μ_x .

Choose $M = X$. Denote $\min_{x \in \emptyset} \mu_x = 1$. For any message in R the action r is chosen.

For a message $(i,j) \in A$, distinguish between two cases:

(i) $\mu_{ij} > 0$

-with probability μ_{ij} the action r is taken.

-with probability $\min_{\{s|is \in R\}} \mu_{is}$ the first aspect is verified.

-with probability $\min_{\{t|tj \in R\}} \mu_{tj}$ the second aspect is verified.

By step 1 $\mu_{ij} + \min_{\{s|is \in R\}} \mu_{is} + \min_{\{t|tj \in R\}} \mu_{tj} = 1$.

(ii) $\mu_{ij} = 0$

Note that $\min_{\{s|is \in R\}} \mu_{is} + \min_{\{t|tj \in R\}} \mu_{tj} \geq 1$. Choose two numbers $p_1 \leq \min_{\{s|is \in R\}} \mu_{is}$ and $p_2 \leq \min_{\{t|tj \in R\}} \mu_{tj}$ satisfying $p_1 + p_2 = 1$. Aspect 1 (2) is verified with probability p_1 (p_2).

Step 3: We will now show that for the mechanism built in Step 2, for every $x \in R$ the probability that the action a is taken is μ_x . Let $x = (i, j)$.

Type (i, j) cannot induce the action a with positive probability unless he sends a message $(i, s^*) \in A$ or $(t^*, j) \in A$. If he utters (i, s^*) the first aspect is verified with probability of at most $\min_{\{s|is \in R\}} \mu_{is} \leq \mu_{ij}$. If he utters (t^*, j) the second aspect is verified with probability of at most $\min_{\{t|tj \in R\}} \mu_{tj} \leq \mu_{ij}$. Thus, (i, j) cannot induce the action a with probability higher than μ_{ij} .

To see that type (i, j) can induce the action a with a probability of exactly μ_{ij} note that if $\mu_{ij} > 0$ then by Step 1 there is an $L, y \in A, z \in R$ and $(i, j) \in R$ such that $\mu_y + \mu_z + \mu_{ij} = 1$. Assume that $y = (i, k)$ and $z = (l, k)$ (the case $y = (k, j)$ can be treated symmetrically). It must be that $\min_{\{s|is \in R\}} \mu_{is} = \mu_{ij}$ since if $\mu_{ij} > \min_{\{s|is \in R\}} \mu_{is} = \mu_{ir}$ for some $r \in X_2$ then $\mu_{ik} + \mu_{ir} + \mu_{lk} < \mu_{ik} + \mu_{ij} + \mu_{lk} = 1$, thus, violating one of the constraints. If (i, j) sends the message (i, k) then:

if $\mu_{ik} > 0$ then the first aspect is verified with probability $\min_{\{s|is \in R\}} \mu_{is} = \mu_{ij}$, and

if $\mu_{ik} = 0$ then the first aspect is verified with probability of at most μ_{ij} and the second aspect is verified with probability of at most $\min_{\{t|tk \in R\}} \mu_{tk} \leq \mu_{lk}$ but since $\mu_{ij} + \mu_{lk} = 1$ the probability that aspect 1 is verified is exactly μ_{ij} .

Step 4: We will show that for the mechanism built in Step 2, for every $x \in A$ the probability that the action a is taken is μ_x . Let $x = (i, j)$.

If $\mu_{ij} = 0$ then type (i, j) will induce the action a with probability 1 by uttering (i, j) .

If $\mu_{ij} > 0$ then if type (i, j) utters (i, j) the action r is taken with probability μ_{ij} . If this type sends another message $(i, k) \in A$ (or $(k, j) \in A$) then he will induce the action a with probability of at most $\min_{\{s|is \in R\}} \mu_{is}$ (or $\min_{\{t|tj \in R\}} \mu_{tj}$). Therefore the mistake probability is at least $1 - \min_{\{s|is \in R\}} \mu_{is}$ (or $1 - \min_{\{t|tj \in R\}} \mu_{tj}$) and by Step 1 (i, j) is a member of some L in which the sum of mistakes is exactly one and thus $1 - \min_{\{s|is \in R\}} \mu_{is} \geq \mu_{ij}$ (or $1 - \min_{\{t|tj \in R\}} \mu_{tj} \geq \mu_{ij}$). ■

Note that in the mechanism constructed in Step 2 of Proposition 1 the listener only verifies whether the speaker's report regarding an aspect is true or not and does not need to observe the exact value of the aspect.

5. A Useful Technique and Some Examples

In some interesting cases finding an optimal mechanism can be done by using a simple technique: finding a mechanism that induces H mistakes and finding H disjoint L 's allows us to conclude that this mechanism is optimal, thus yielding a mistake probability of $H/|X|$.

The following examples are used to demonstrate this technique. The examples also give us some intuition as to when optimality can be obtained by deterministic mechanisms and when it requires that the listener use randomization to determine the aspect to be checked.

Example 2

Consider the problem (X, A) where $X_1 = X_2 = \{1, \dots, 5\}$ and $A = \{x | x_1 + x_2 \geq 7\}$. In Figure 2, each entry stands for an element in X and the elements in A are indicated by the letter A . We denote 5 disjoint L 's (the three elements of each L are indicated by the same number):

1	A1	A	A	A
2	5	A2	A5	A
3			A3	A
4			5	A4
	1	2	3	4

Figure 2

Following is a direct mechanism that induces 5 mistakes and thus is optimal: For any message m such that $m_k \leq 4$ for both k , the action r is taken. Otherwise, an aspect k for which $m_k > 4$ is verified. In fact, this mechanism amounts to simply asking the speaker to present an aspect with a value of 5. The five mistakes are with respect to the three types in A which do not contain a component of 5 and the two types in R that contain a component of 5.

Examples 1 and 2 can be generalized (see Proposition 3) to yield the conclusion by which the optimal mechanism does not require randomization when the speaker is aiming to persuade the listener that the average of the two values of the two aspects is above a certain threshold.

Example 3

Although we confine our analysis to the case of two aspects, this example demonstrates how the L -principle can be applied to solve for optimal mechanisms when the number of aspects is

greater than 2. We will also show that the conclusion reached above, i.e. that randomization is not needed for the case in which the speaker tries to persuade the listener that the average of the values of the two aspects is above a certain threshold, does not hold for the case in which the number of aspects is greater than 2.

Consider the problem where $n = 3$, $X_k = \{0, 1\}$ for $k = 1, 2, 3$ and $A = \{(x_1, x_2, x_3) | \sum_k x_k \geq 2\}$. An optimal mechanism is to ask the speaker to report two aspects which received the value 1 and to verify each of them with probability 1/2. This mechanism yields 1.5 mistakes (mistake probability of 3/16) since only the three types (1,0,0), (0,1,0) and (0,0,1) can mislead the listener with probability 1/2.

To see that this is an optimal mechanism we will apply the L -principle. For any mechanism the following three inequalities must hold:

$$\mu_{(1,1,0)} + \mu_{(1,0,0)} + \mu_{(0,1,0)} \geq 1$$

$$\mu_{(1,0,1)} + \mu_{(1,0,0)} + \mu_{(0,0,1)} \geq 1$$

$$\mu_{(0,1,1)} + \mu_{(0,1,0)} + \mu_{(0,0,1)} \geq 1$$

Let $e(x) = x_1 + x_2 + x_3$. The minimum of $\sum_{x \in X} \mu_x$ subject to the following constraint $\sum_{e(x)=2} \mu_x + 2 \sum_{e(x)=1} \mu_x \geq 3$, implied by summing up the three inequalities, is attained when $\mu_x = 1/2$ for all $e(x) = 1$ and $\mu_x = 0$ for any other x . Thus, the number of mistakes cannot fall below 1.5.

Within the set of deterministic mechanisms an optimal mechanism is to take the action a iff the speaker can show that either aspect 1 or aspect 2 has the value 1. This mechanism induces two mistakes one with regard to type (1,0,0) and the other in regard to type (0,1,0).

Example 4

Assume that the speaker tries to persuade the listener that the values of the two aspects are

distinct. This would fit a situation in which, for example, the speaker claims that he has taken two different actions on two different occasions but the listener can verify no more than one of them. Formally, consider the problem (X, A) where $k = \{1, 2, 3, \dots, I\}$ and $A = \{(x_1, x_2) | x_1 \neq x_2\}$. The optimal mechanism will be shown to be non-deterministic in this case.

Intuitively, any information about only one of the aspects provides no useful information to the listener and thus asking the speaker to state the value of the two aspects and randomly verifying one of them might be better for the listener than any deterministic mechanism.

Figure 3a presents the problem for $I = 4$ by indicating the elements in the set R . Two disjoint L 's are denoted by the numbers 1 and 2.

		1	$R1$
		$R1$	
2	$R2$		
$R2$			

Figure 3a

*	*	$R *$
*	$R *$	
$R *$		

Figure 3b

Figure 3b presents the problem for $I = 3$. The maximal number of disjoint L 's is one. Notice the six starred elements - three in A and three in R . They produce three (non-disjoint) L 's. Any mechanism induces mistake probabilities $(\mu_x)_{x \in X}$ satisfying:

$$\mu_{1,3} + \mu_{1,1} + \mu_{3,3} \geq 1$$

$$\mu_{1,2} + \mu_{1,1} + \mu_{2,2} \geq 1$$

$$\mu_{2,3} + \mu_{2,2} + \mu_{3,3} \geq 1$$

which imply that the sum of mistakes with respect to these six elements must be at least 1.5.

Generalizing these two examples we can see that the minimal number of mistakes is $I/2$. The fair random mechanism (such that by following a message in A each of the two aspects is

verified with probability 0.5) induces $I/2$ mistakes and is thus optimal.

Any deterministic mechanism induces a vector $(\mu_x)_{x \in X}$ of mistakes with $\mu_x \in \{0, 1\}$ for all x . If there is an i for which $\mu_{ii} = 0$, then for any $j \neq i$ either $\mu_{jj} = 1$ or $\mu_{ij} = 1$ since if $\mu_{jj} = 0$ the constraint $\mu_{ij} + \mu_{ii} + \mu_{jj} \geq 1$ implies $\mu_{ij} = 1$. Therefore $\sum_{x \in X} \mu_x \geq I - 1$. Thus, any deterministic mechanism induces at least $I - 1$ mistakes.

In a deterministic mechanism that induces $I - 1$ mistakes the speaker is asked to present an aspect whose realization is not 1 (thus yielding mistakes only for the types (i, i) with $i \neq 1$).

Note that in the mirror problem in which the listener tries to persuade the listener the listener that the two aspects received the same value it is also true that any information about one of the aspects provides no useful information to the listener. However, this is a degenerate case and the optimal mechanism here is to choose r independently of the speaker's message without the need to check any of the aspects. Assume, $X_1 = X_2 = \{1, \dots, I\}$ and $A = \{x | (x_1, x_2) | x_1 = x_2\}$ and $I > 2$. The constant r mechanism in this case induces I mistakes. To see that one cannot further reduce the number of mistakes note that the I sets $\{(i, i), (i + 1, i), (i, i + 1)\}$ for $i = 1, \dots, I - 1$ and $\{(I, I), (1, I), (I, 1)\}$ consist of a collection of disjoint L 's.

Example 5

Up to now, all optimal mechanisms have had the property that the induced probability of mistake with respect to a type in R was 0, 1 or 0.5 whereas for types in A the probability of mistake was either 0 or 1. In the following example any optimal mechanism induces a mistake probability of 0.5 for at least one type in A .

Figure 4a presents the problem (X, A) , where $X_1 = \{1, \dots, 8\}$, $X_2 = \{1, \dots, 7\}$ and the types in A are denoted by A .

*	*	*	*	*	A *		
*	*	*	*	*	A *		
*	A	A	A	A	*	*	*
A	*	A	A	A	*	*	*
A	A	*	A	A	*	*	*
A	A	A	*	A	*	*	*
A	A	A	A	*	*	*	*

Figure 4a

9	10	1	2	5	15A	15	
8	6	3	4	7	16A	16	
11	A	1A	2A	14A	14	1	2
11A	11	3A	4A	A	15	3	4
A	6A	13	12A	5A	12	5	6
8A	A	13A	12	7A	13	7	8
9A	10A	A	A	14	16	9	10

Figure 4b

Figure 4b presents 16 disjoint L 's: each L is indicated by a number. A solution to $P(X,A)$ is $\mu_x = 1/2$ for any of the 32 types indicated by a star in Figure 4a and $\mu_x = 0$ otherwise. Note that $\mu_{66} = \mu_{67} = 1/2$ although $(6,6)$ and $(6,7)$ are in A .

Consider $(\mu_x)_{x \in X}$, which is a solution for $P(X,A)$. We will show that either μ_{66} or μ_{67} is not an integer.

First, note that $\mu_{6,6} + \mu_{6,7} \leq 1$. Notice that in Figure 4c we indicate 15 disjoint L 's that do not contain any of the elements in the box $\{6,7,8\} \times \{6,7\}$ and thus the sum of mistakes in that box cannot exceed 1.

Note also that the 16 L 's in Figure 6d do not contain $(8,6)$ and $(8,7)$ and thus $\mu_{86} = \mu_{87} = 0$. Similarly, $\mu_{76} = \mu_{77} = 0$ (we obtain an alternative collection of 16 disjoint L 's by replacing $(7,6)$ with $(8,6)$ and $(7,7)$ with $(8,7)$ in Figure 4b).

9	10	1	2	5	A		
8	6	3	4	7	A		
13	11A	1A	2A	A	11	1	2
A	11	3A	4A	12A	12	3	4
13A	6A	14	A	5A	13	5	6
8A	A	14A	15	7A	14	7	8
9A	10A	A	15A	12	15	9	10

Figure 4c

Now assume that both $\mu_{6,6}$ and $\mu_{6,7}$ are integers. Then there is $j \in \{6, 7\}$ so that $\mu_{6,j} = 0$. For any $i = 1, \dots, 5$ it must be $\mu_{6,j} + \mu_{6,i} + \mu_{7j} \geq 1$ and thus $\mu_{6,i} = 1$. However, none of the 12 disjoint L 's in Figure 4d contain any of the 5 types $(6, i)$ where $i = 1, \dots, 5$ and hence the total number of mistakes is at least 17 which is a contradiction !

9	10	1	2	5	A		
8	6	3	4	7	A		
11	11A	1A	2A	A		1	2
A	11	3A	4A	A		3	4
A	6A	12	12A	5A		5	6
8A	A	A	12	7A		7	8
9A	10A	A	A			9	10

Figure 4d

Comment (The Dual Problem): Let (X, A) be a problem and let $T(X, A)$ be the set of all its L 's. $P(X, A)$ is the linear programming optimization. The dual problem $D(X, A)$ is:

$$\max \sum_{\Delta \in T(X, A)} \lambda_{\Delta}$$

$$\text{s.t. } \sum_{x \in \Delta \in T(X, A)} \lambda_{\Delta} \leq 1 \text{ for all } x \in X$$

$$\text{and } 0 \leq \lambda_{\Delta} \text{ for all } \Delta \in T.$$

Finding a collection of disjoint L 's is equivalent to finding a point within the constraints of $D(X, A)$ for which $\lambda_{\Delta} = 1$ for any Δ in the collection and $\lambda_{\Delta} = 0$ for Δ not in the collection. We know that the values of the solutions of $P(X, A)$ and $D(X, A)$ must be identical. Thus, finding a number of disjoint L 's and a mechanism that induces the same number of mistakes, is the same as finding two vectors, one satisfying the constraints of $P(X, A)$ and another satisfying the constraints of $D(X, A)$, that yield the same value for the two objective functions.

Note for example the case of $I = 3$ in Example 4. Assigning $\lambda_{\Delta} = 1/2$ for the three L 's

$\{(1, 3), (1, 1), (3, 3)\}$, $\{(1, 2), (1, 1), (2, 2)\}$ and $\{(2, 3), (2, 2), (3, 3)\}$ and $\lambda_\Delta = 0$ otherwise we obtain a value of 1.5 for the dual problem. This is exactly the value of the primal problem where we assign $\mu_x = 1/2$ to the three points on the main diagonal and $\mu_x = 0$ otherwise.

6. The Simple Structure of Optimal Mechanisms

We have seen several examples in which the optimal mechanisms only used a form of randomization that could be carried out by tossing a fair coin. We will now show that for $n = 2$ we can always construct an optimal mechanism without applying other forms of randomization.

Proposition 2: For every finite problem (X, A) there exists an optimal direct mechanism $(M = X)$ such that if $m \in R$ the listener takes the action r whereas if $m \in A$ the listener does one of the following:

- (i) takes the action r .
- (ii) takes the action r with probability $1/2$ and verifies one aspect with probability $1/2$.
- (iii) verifies each aspect with probability $1/2$.
- (iv) verifies one aspect with probability 1.

(Using our notation, for every $m \in R$, $f(m)$ is the degenerate lottery r and for every message $m = (m_1, m_2) \in A$, $f(m)$ is a lottery $\langle \pi_0, d_0; \pi_1, d_1; \pi_2, d_2 \rangle$ where all π_i are in $\{0, 1/2, 1\}$, $d_0 = r$, $d_1(x_1) = a$ iff $x_1 = m_1$ and $d_2(x_2) = a$ iff $x_2 = m_2$.)

Proof: Let $(\mu_x)_{x \in X}$ be a solution to $P(X, A)$. A proposition due to Noga Alon (see Alon (2003)) states that if $(\alpha_x)_{x \in X}$ is an extreme point of the set of all vectors satisfying the constraints

in $P(X,A)$, then $\alpha_x \in \{0, 1/2, 1\}$ for all $x \in X$. As a solution of a linear programming problem the vector $(\mu_x)_{x \in X}$ is an extreme point and thus $\mu_x \in \{0, 1/2, 1\}$ for all $x \in X$. The construction of an optimal mechanism in Proposition 1 implies the rest of our claim since for every $i \in X_1$ and $j \in X_2$ the numbers $\min_{\{s|is \in R\}} \mu_{is}$ and $\min_{\{t|tj \in R\}} \mu_{tj}$ are all within $\{0, 1/2, 1\}$. ■

Comment: Our initial conjecture was that if $(\alpha_x)_{x \in X}$ is an extreme point of the set of all vectors satisfying the constraints in $P(X,A)$, then $\alpha_x \in \{0, 1\}$ for all $x \in A$ and $\alpha_x \in \{0, 1/2, 1\}$ for all $x \in R$. Noga Alon showed that we were only partially right and proved the modification of our conjecture used in the proof above.

Remark: We present our analysis for the case of uniform probability measure over the set X . Proposition 2 can be extended to any probability measure. Denote by $p(x)$ the probability of type x and consider the modified problem:

$$\min \sum_{x \in X} p(x) \mu_x$$

$$\text{s.t. } \mu_{ij} + \mu_{is} + \mu_{tj} \geq 1 \text{ for all } (i,j) \in A, (i,s) \in R \text{ and } (t,j) \in R$$

$$\text{and } 0 \leq \mu_x \text{ for all } x \in X$$

This linear problem is different from $P(X,A)$ only in the objective function. It is easy to see that Proposition 1 is valid for the more general case as well, and since the extreme points of the set of vectors satisfying the constraints are unchanged, Proposition 2 can also be generalized to non-uniform probability measures.

7. A Sufficient Condition for the Optimality of Deterministic Mechanisms

We are now interested in identifying conditions under which the optimal mechanisms are deterministic. The following proposition refers to the case in which the set of types is a continuum although it also gives some insight into the finite case as well.

Given a problem (X, A) where $X = [0, 1] \times [0, 1]$ define a partial ordering $s >_1 s'$ if for every $t \in [0, 1]$ “ $(s', t) \in A$ implies $(s, t) \in A$ ”. The meaning of $s >_1 s'$ is that the information $x_1 = s$ is a better indication that $x \in A$ than the information that $x_1 = s'$. Similarly, define the partial ordering $>_2$. We say that a set A is *monotonic* if $s >_k s'$ iff $s > s'$ for $k = 1, 2$. For example, the set A in Example 1 is monotonic.

Proposition 3: Let $X = [0, 1] \times [0, 1]$ and assume that A is monotonic and that R is closed and convex. Then there exists an optimal mechanism which is deterministic.

Proof: Let $E_1 = \max\{x_1 | (x_1, 0) \in R\}$ and $E_2 = \max\{x_2 | (0, x_2) \in R\}$.

Let $\varphi_2(x_1) = \max\{x_2 | (x_1, x_2) \in R\}$ and $\varphi_1(x_2) = \max\{x_1 | (x_1, x_2) \in R\}$. By definition $\varphi_1(E_2) = 0$ and $\varphi_2(E_1) = 0$. Since R is closed and A is monotonic, the function $(\varphi_1(x_2), \varphi_2(x_1))$ is well defined and continuous from $[0, E_1] \times [0, E_2]$ into itself.

The equation $\varphi_2(2z_1) = z_2$ is a continuous and decreasing curve which contains the points $(0, E_2)$ and $(E_1/2, 0)$ and the equation $\varphi_1(2z_2) = z_1$ is a continuous and decreasing curve which contains the points $(E_1, 0)$ and $(0, E_2/2)$. Thus, there is a pair (α, β) in the box $[0, E_1/2] \times [0, E_2/2]$, satisfying $\varphi_2(2\alpha) = \beta$ and $\varphi_1(2\beta) = \alpha$.

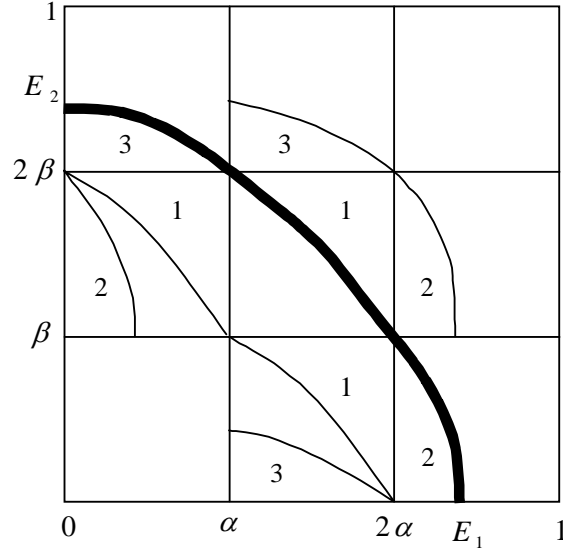


Figure 5

Figure 5 extends the idea embedded in Example 1. The set R is the area below the bold curve. The mistake probability is at least the size of the three “triangles”:

$$T_1 = \{(x_1, x_2) \in A | x_1 \leq 2\alpha \text{ and } x_2 \leq 2\beta\}$$

$$T_2 = \{(x_1, x_2) \notin A | x_1 > 2\alpha\}$$

$$T_3 = \{(x_1, x_2) \notin A | x_2 > 2\beta\}.$$

Note that by the convexity of R the sets $\{(x_1 - \alpha, x_2) | (x_1, x_2) \in T_1\}$ and $\{(x_1 - 2\alpha, x_2 + \beta) | (x_1, x_2) \in T_2\}$ have an empty intersection. Similarly, $\{(x_1, x_2 - \beta) | (x_1, x_2) \in T_1\}$ and $\{(x_1 + \alpha, x_2 - 2\beta) | (x_1, x_2) \in T_3\}$ are disjoint.

Finally, consider the direct mechanism where by following a message (m_1, m_2) aspect 1 is verified if $m_1 \geq 2\alpha$, aspect 2 is verified if $m_2 \geq 2\beta$ and the action r is taken otherwise. The mistake probability induced by this mechanism equals exactly to the sum of the sizes of T_1, T_2 and T_3 . ■

Note the rationale of the optimal deterministic mechanism described in the proof. The speaker can persuade the listener to take the action a by showing that the first aspect received a value above 2α or that the second aspect received a value above 2β . If the speaker refrains from

taking this route the listener will conclude that the speaker's type x satisfies $x_1 \leq 2\alpha$ and $x_2 \leq 2\beta$. When the listener finds out that the first aspect received the value x_1 he compares between $prob\{x_2 \leq 2\beta | (x_1, x_2) \in A\}$ and $prob\{x_2 \leq 2\beta | (x_1, x_2) \in R\}$. The number 2α has the property that if $x_1 > 2\alpha$, it is more likely that $x \in A$ and if $x_1 < 2\alpha$, it is more likely that $x \in R$.

The following example is finite, but we bring it here as an hint that when the problem is monotonic but the set R is not convex, the optimal mechanism may not be deterministic.

Example 6

Let $X_1 = X_2 = \{1, 2, \dots, 6\}$. The elements in R are indicated in Figure 6.

R^*	*			*	
$R1$			1		
$R3$	3				
$R2$	R^*	2		*	
R	$R3$	$R2$	$R1$	R^*	

Figure 6

The maximal number of disjoint L 's is 4 although the optimal mechanism yields 4.5 mistakes.

Notice the six starred elements, three in A and three in R . They produce three (non-disjoint) L 's. Any mechanism induces mistake probabilities $(\mu_x)_{x \in X}$ satisfying:

$$\mu_{5,5} + \mu_{1,5} + \mu_{5,1} \geq 1$$

$$\mu_{2,5} + \mu_{2,2} + \mu_{1,5} \geq 1$$

$$\mu_{5,2} + \mu_{2,2} + \mu_{5,1} \geq 1$$

which imply that the sum of mistakes with respect to these six elements must be at least 1.5.

At least three additional mistakes must be induced with respect to the three disjoint L 's indicated

by the numbers 1, 2 and 3 in Figure 4. Thus, any mechanism must induce at least 4.5 mistakes.

The following direct mechanism yields 4.5 mistakes: if $m \in R$ the action r is chosen; if the speaker claims that one of the aspects received the value 6, then this aspect is verified; for any other message in A , the aspect to be verified is determined by a fair coin toss. Under this mechanism, the 9 types in $R - \{(1, 1)\}$ induce the action a with probability 0.5. Within the set of deterministic mechanisms, an optimal mechanism, yielding 5 mistakes, is to ask the speaker to show that either $x_1 \geq 6$ or $x_2 \geq 2$.

8. Optimal Deterministic Mechanisms

In this section we study optimality within the class of deterministic mechanisms. We can think about a deterministic mechanism in the following way: once the speaker has uttered a message m , the listener checks one aspect $k(m)$ with probability 1 and chooses a iff the value of the aspect is in some set $V(m) \subseteq X_{k(m)}$. A speaker of type (x_1, x_2) will be able to induce the listener to take the action a if there is a message m such that $x_{k(m)} \in V(m)$. Denote $V_k = \cup_{k(m)=k} V(m)$. Thus, a type (x_1, x_2) will induce a if and only if $x_k \in V_k$ for at least one k . To summarize, for any deterministic mechanism there are two sets $V_1 \subseteq X_1$ and $V_2 \subseteq X_2$ such that the probability of a mistake is the probability of $\{(x_1, x_2) \in A \mid \text{for no } k, x_k \in V_k\} \cup \{(x_1, x_2) \in R \mid \text{for at least one } k, x_k \in V_k\}$. We call the sets V_1 and V_2 the *sets of persuasive facts*.

We are now able to derive a simple necessary condition for a mechanism to be optimal within the set of deterministic mechanisms:

Proposition 4: Let (X, A) be a finite problem. For a mechanism to be optimal within the set of deterministic mechanisms, its persuasive facts $V_1 \subseteq X_1$ and $V_2 \subseteq X_2$ must satisfy:

for any $x_1 \in V_1$ $prob\{x_2 \notin V_2 | (x_1, x_2) \in A\} \geq prob\{x_2 \notin V_2 | (x_1, x_2) \in R\}$ and

for any $x_1 \notin V_1$ $prob\{x_2 \notin V_2 | (x_1, x_2) \in A\} \leq prob\{x_2 \notin V_2 | (x_1, x_2) \in R\}$

(the term *prob* refers to the uniform probability on X_2). Similar condition holds for V_2 .

Proof: Assume, for example, that $s \in V_1$ but that

$prob\{x_2 \notin V_2 | (s, x_2) \in A\} < prob\{x_2 \notin V_2 | (s, x_2) \in R\}$. Eliminating s from V_1 will decrease the mistake probability. To see this, note first that every type x such that either $x_1 \neq s$ or $x_2 \in V_2$ can induce the action a iff he could induce it prior to the elimination of s from V_1 . Any type x such that $x_1 = s$ and $x_2 \notin V_2$ could induce the action a prior to the elimination but cannot do so after the elimination. Since $prob\{x_2 \notin V_2 | (s, x_2) \in A\} < prob\{x_2 \notin V_2 | (s, x_2) \in R\}$ elimination of such an s reduces the mistake probability. ■

It follows that for an optimal deterministic mechanism, if $s >_k t$ then if $t \in V_k$ so is s . It also follows that when A is monotonic, any mechanism which is optimal among the deterministic mechanisms, is characterized by two cutting points z_1 and z_2 such that the listener chooses a only once he has verified that one of the aspects k received a value above z_k .

Proposition 4 is stated for finite problems. It is not difficult to extend it to the infinite case taking into account measurability conditions.

Note that the condition stated in Proposition 4 is necessary but not sufficient for a mechanism to be optimal within the set of deterministic mechanisms: Returning to example 3 with $X_1 = X_2 = \{1, 2, 3, 4\}$, a mechanism with $V_1 = V_2 = \{3, 4\}$ satisfies the conditions in the proposition and yields 4 mistakes, while the mechanism with $V_1 = V_2 = \{2, 3, 4\}$ yields only 3 mistakes.

The section ends with another example which demonstrates that randomization might also be necessary for non-finite problems.

Example 7

Suppose that a child wishes to persuade one of his parents that overall he has allocated his efforts fairly evenly between studying and leisure during the last two days but the parent can only verify this for at most one of the two days. Denote by $x_i \in X_k = [0, 1]$ the fraction of day i that the child has devoted to studying and let $A = \{(x_1, x_2) \mid 1/2 \leq x_1 + x_2 \leq 3/2\}$.

The fair random mechanism induces probability of mistake of $1/8$ (half the types in R). We will show that any deterministic mechanism will have probability of mistake of at least $1/6$.

Recall that any deterministic mechanism is characterized by the sets of persuasive facts V_1 and V_2 . If $\text{prob}\{x_2 \notin V_2 \mid (0, x_2) \in A\} \geq \text{prob}\{x_2 \notin V_2 \mid (0, x_2) \in R\}$ then $\text{prob}\{x_2 \notin V_2 \mid (x_1, x_2) \in A\} \geq \text{prob}\{x_2 \notin V_2 \mid (x_1, x_2) \in R\}$ for all $x_1 \leq x_1^*$ for some x_1^* and thus extending V_1 to the set $[0, x_1^*]$ will not increase the probability of mistake. In the modified mechanism we have $\text{prob}\{x_1 \notin V_1 \mid (x_1, 0) \in A\} \geq \text{prob}\{x_1 \notin V_1 \mid (x_1, 0) \in R\}$ and thus also $\text{prob}\{x_1 \notin V_1 \mid (x_1, x_2) \in A\} \geq \text{prob}\{x_1 \notin V_1 \mid (x_1, x_2) \in R\}$ for all $x_2 \leq x_2^*$ for some x_2^* and thus we can modify the mechanism so that $V_2 = [0, x_2^*]$ without increasing the mistake probability. In the class of deterministic mechanisms with $V_k = [0, x_k^*]$ having $x_1^* = x_2^* = 2/3$ minimizes the probability of mistake and yields a mistake probability of $1/6$.

If $\text{prob}\{x_2 \notin V_2 \mid (0, x_2) \in A\} < \text{prob}\{x_2 \notin V_2 \mid (0, x_2) \in R\}$ then since $(0, x_2) \in A$ if and only if $(1, x_2) \in R$ (except at the single point $x_2 = 1/2$), we have $\text{prob}\{x_2 \notin V_2 \mid (1, x_2) \in A\} > \text{prob}\{x_2 \notin V_2 \mid (1, x_2) \in R\}$ and analogous to the above the choice $V_1 = [1/3, 1]$ and $V_2 = [1/3, 1]$ minimizes the probability of mistake and yields the mistake probability of $1/6$.

9. Related Literature

The model in this paper can be thought of as a principal agent problem. Some previous papers have studied principal agent problems in situations where the agent is able to show some “hard evidence”. The literature we are aware of has studied the circumstances under which the revelation principle holds, whereas our main interest is in characterizing the optimal

mechanisms.

Green and Laffont (1986) studies mechanisms in which the set of messages each type can send can depend on the type and is a subset of the set of types. Their framework does not allow the listener to randomize. Furthermore, their model does not cover the case in which the speaker can show the value of the realization of one of the aspects. In particular, assuming in their framework that a type (i,j) can send only messages like (i,s) or (t,j) is not the same as assuming that he can present one of the aspects. The reason is that a message (m_1, m_2) would not reveal whether the agent actually showed that the realization of aspect 1 is m_1 or that he showed that the realization of aspect 2 is m_2 .

In Bull and Watson (2002) an agent can also show evidence. A key condition in their paper is what they call “normality”: if a type x can distinguish himself from type x' and from x'' then he can distinguish himself from both, a condition which does not hold in our framework. They do not consider randomized mechanisms.

A related paper is Fishman and Hagerty (1990). According to one interpretation what they do is to analyze the optimal deterministic mechanisms for the problem $(\{0, 1\}^n, \{x | \sum_k x_k > b\})$ for some b .

Our interest in this paper is rooted in Glazer and Rubinstein (2001) where we study the design of optimal deterministic debate mechanisms in a specific example. (Other models of optimal design of debate rules with hard evidence are Shin (1994) and Lipman and Seppi (1995).) The two models are quite different but nevertheless have some common features. In both models there is a listener and speaker(s); the listener has to take an action after listening to arguments made by the speaker(s); an instance is characterized by the realization of several aspects and the speaker(s) knows the realization of the aspects while the listener does not; the speaker(s) can be asked to present “hard” evidence showing the realization of an aspect; the listener is constrained by the number of aspects he can check and, hence, has to base his decision on only partial information. In both papers we look for a mechanism that minimizes the

probability that the listener will take the wrong action.

References

- Alon, Noga (2003) A Comment on Extreme Points (memo).
- Bull, Jesse and Joel Watson (2002), Hard Evidence and Mechanism Design, working paper.
- Fishman, Michael J. and Kathleen M. Hagerty (1990), The Optimal Amount of Discretion to Allow in Disclosures, *Quarterly Journal of Economics*, 105, 427-444.
- Glazer, Jacob and Ariel Rubinstein (2001), Debates and Decisions, On a Rationale of Argumentation Rules, *Games and Economic Behavior*, 36 (2001), 158-173
- Green, Jerry and Jean-Jacques Laffont (1986), Partially Verifiable Information and Mechanism Design, *Review of Economic Studies*, 447-456.
- Lipman, Barton L. and Duane J. Seppi, (1995), Robust Inference in Communication Games with Partial Provanility, *Journal of Economic Theory*, 66, 370-405.
- Shin, Hyun Song (1994), The Burden of proof in a Game of Persuasion, *Journal of Economic Theory*, 64, 253-264.